Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Identification of novel genes associated with dysregulation of B cells in patients with primary Sjögren's syndrome

Jun Inamo
Keio University School of Medicine https://orcid.org/0000-0002-9927-7936
Katsuya Suzuki
Keio Gijuku Daigaku Rikogakubu Daigakuin Rikogaku Kenkyuka
Masaru Takeshita
Keio University School of Medicine
Yoshiaki Kassai
Takeda Pharamaceutical Company Limited
Maiko Takiguchi
Takeda Pharamaceutical Company Limited.
Rina Kurisu
Takeda Pharamaceutical Company Limited.
Yuumi Okuzono
Takeda Pharamaceutical Company Limited.
Shinya Tasaki
Takeda Pharamaceutical Company Limited.
Akihiko Yoshimura
Keio University School of Medicine
Tsutomu Takeuchi (tsutake@z5.keio.jp)
https://orcid.org/0000-0003-1111-8218

Research article

Keywords: primary Sjögren's syndrome, B cells, long non-coding RNA, transcriptome, gene co-expression network

Posted Date: June 5th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-26390/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Version of Record: A version of this preprint was published on June 22nd, 2020. See the published version at https://doi.org/10.1186/s13075-020-02248-2.

Abstract

Background. The aim of this study was to identify the molecular mechanism of dysregulation of B cell subpopulations of primary Sjögren's syndrome (pSS) at the transcriptome level.

Methods. We enrolled patients with pSS (n=6) and healthy controls (HC) (n=6) in the discovery cohort using microarray and pSS (n=14) and HC (n=12) in the validation cohort using quantitative PCR (qPCR). Peripheral B cells acquired from these subjects were separated by cell sorting into four subsets: $CD38^{-}IgD^{+}$ (Bm1), $CD38^{+}IgD^{+}$ (naïve B cells), $CD38^{high}IgD^{+}$ (pre-germinal centre B cells) and $CD38^{\pm}IgD^{-}$ (memory B cells). We performed differentially expressed genes (DEGs) analysis and weighted gene co-expression network analysis (WGCNA).

Results. Expression of the long non-coding RNA *LINC00487* was significantly upregulated in all B cell subsets, as was that of HLA and interferon (IFN) signature genes. Moreover, the normalized intensity value of *LINC00487* significantly correlated with the disease activity score of all pSS B cell subsets. Studies of human B cell lines revealed that the expression of *LINC00487* was strongly induced by IFNα. WGCNA revealed six gene clusters associated with the B cell subpopulation of pSS. Further, *SOX4* was identified as an inter-module hub gene.

Conclusion. Our transcriptome analysis revealed key genes involved in the dysregulation of B cell subpopulations associated with pSS.

Trial registration: Not required.

Background

Primary Sjögren’s syndrome (pSS) is a systemic autoimmunedisease characterized by exocrine gland dysfunction, which leads todryness of the eyes and mouth [1]. Upregulation of the interferon(IFN) pathway [2] and dysregulation of B cells play a critical rolein the pathogenesis of pSS. First, the loss of B cell tolerance canlead to overproduction of anti-Sjögren’s syndrome-related antigen A(anti-SSA) autoantibodies. Autoreactive B cells are important inthe development of clinical disease, because serum autoantibodiesmay precede the onset of dryness. Second, IFN-stimulated genes(ISGs) are upregulated in B cells in peripheral blood as well as inthe salivary gland lesions of patients with pSS [3,4]. Third, IFNαpromotes loss of tolerance and development of autoreactive B cells[5]. Genome-wide association studies identified single nucleotidepolymorphisms of ISGs that are associated with the risk of pSS[6,7]. Further, viral infection, which enhances the IFN pathway,may contribute to the pathogenesis of pSS [8].

Non-coding RNAs, which are emerging as critical regulators ofsignal transduction pathways, may be involved in the immune system.Non-protein coding microRNAs are involved in the dysregulation in Bcells [9] and in the upregulation of the IFN pathway in patientswith pSS [10]. Long non-coding RNAs (lncRNAs) are involved indiverse regulatory functions, including the modulation of chromatinstructure and post-transcriptional regulation affecting thestability of mRNAs and proteins [11]. Recent evidence

indicatesthat lncRNAs, which have become the focus of studies on autoimmunediseases, may contribute to the pathogenesis of pSS throughmultiple signal transduction pathways [12]. For example,differentially expressed lncRNAs in the labial salivary glands ofpSS patients were identified using microarray analysis [13]. Whilemulti-omics studies of the whole-blood transcriptome highlight theimportance of cytotoxic CD8 T cells in the pathogenesis of pSS, theB cell subpopulation is associated with clinical traits [14].

Considering that gene expression pattern is tissue-specific[15], it is better to focus on the target cell subpopulation toavoid noise from other cells in transcriptome analysis. However, weare unaware of studies that focus on the effects of dysregulationof transcriptomes, including lncRNAs, of B cells on thepathogenesis of pSS. Serum level of B cell activating factor wasassociated with disease activity of pSS [16-18]. In addition, thedistribution of peripheral B cell subsets is profoundly altered inpatients with pSS [19]. Staining for IgD/CD38 is helpful forstudying circulating B cell subsets, separating developmentalstages from naive to memory B cells (Bm1 to Bm5) [20]. Using thisgating strategy, we previously reported that the proportion ofpre-GC B cells ($IgD^{+}/CD38^{high}$) was higher inpatients with pSS compared with healthy controls and associatedwith clinical traits, such as disease activity [19], indicating theimportance of evaluating the involvement of each B cell subset inthe pathogenesis of pSS.

To determine the role of the B cell subset in the pathogenesisof pSS, we investigated potentially significant genes according totheir differential levels of expression and connectivity with othergenes. Thus, the current study consists of two sections:differentially expressed genes (DEGs) analysis and weighted geneco-expression network analysis (WGCNA). First, we describe theupregulation of the interferon signalling pathway and thedifferential expression of genes encoding human leukocyte antigen(HLA) molecules and *LINC00487* in B cell subpopulations ofpatients with pSS compared with healthy controls (HC). Theexpression levels of *LINC00487* correlated with the diseaseactivity of pSS and IFN-signature genes; and *LINC00487* wasinduced by IFNα. Second, using WGCNA, we identified genes ofco-expression networks specific to a B cell subset of patients withpSS, suggesting that aberrant molecular interactions in B cellscontribute to the aetiology of pSS.

Methods

Patients and controls

The study protocol is shown in Figure 1A. We enrolled patientswith pSS (n=6) and HC (n=6) matched for age and sex (SupplementaryTable 1). The patients with pSS fulfilled the 2002American-European criteria for SS [21] and the 2012 AmericanCollege of Rheumatology classification criteria for SS [22]. Allpatients were female and had symptom of dryness, anti-SSAantibodies and biopsy-proven sialadenitis. The patients had notreceived immunosuppressive therapy, and the HCs did not haveimmunological disorders. Clinical and serological information ofpatients with pSS were collected from their medical records. Thedisease activity of pSS was assessed according to the guidelines ofthe European League against Rheumatism and the Primary Sjögren'ssyndrome disease activity index (ESSDAI) [23, 24]. The

distributionof ESSDAI domain and clinical information of individual patientswere described in Supplementary Table 17 and Supplementary Table18.

This study was performed in accordance with relevant guidelinesand regulations. The Ethics Committee of Keio University School ofMedicine approved this study (IRB No. 20110258), and writteninformed consent was obtained from each subject before bloodcollection.

Cell sorting

Peripheral blood mononuclear cells from patients with pSS and HCwere separated using gradient centrifugation with Lymphoprep(Axis-Shield; Oslo, Norway). Gating strategy was shown inSupplementary Figure 1. Peripheral $CD19^+$ B cells wereprepared with anti-CD19 antibody-coated microbeads (MiltenyiBiotec). As previously reported [19], the peripheral$CD19^+$ B cells were incubated with anti-IgD and CD38antibodies for fluorescence-activated cell sorting (FACS) analysis(FACSAria III flow cytometer, BD Biosciences). We defined subsetsof B cells as follows: Bm1 cells; $CD38^-IgD^+$,naïve B cells; $CD38^+IgD^+$, pre-germinal centre(pre-GC) B cells; $CD38^{high}IgD^+$ and memory Bcells; $CD38^{\pm}IgD^-$.

DEGs analysis

Total RNA was extracted from B cell subsets and transcribed intocDNA using NucleoSpin RNA (Macherey Nagel) and ReverTra Ace qPCR RTMaster Mix (Toyobo). Gene expression was measured using the HumanGenome U133 Plus 2.0 Array (Affymetrix). We applied percentileshift normalization to the raw signal data acquired from amicroarray and annotated each probe with its gene symbol usingGeneSpring software (Agilent Technologies). Probes withinterquartile ranges in the lowest 20% were excluded. We nextselected probes with >2.0-changes for pSS vs HC in any one Bcell subset to identify DEGs. We controlled for the false discoveryrate using the Bonferroni multiple testing-corrected p value<0.05. To functionally characterize DEGs identified in each Bcell subpopulation from microarray analysis, we performed pathwayanalysis using Enrichr's plugin [25] BioPlanet [26]. BioPlanetdatabase incorporates more than 1500 human pathways sourced frompublicly available, manually curated sources. In pathway analysis,p-value was adjusted using the Benjamini-Hochberg method forcorrection for multiple hypotheses testing.

WGCNA

To explore novel gene co-expression networks and common hubgenes, we created another gene set. In order to select genes thatare steadily expressed in each B cell subpopulation of pSS and HC(totally, 8 subpopulations; Bm1, naïve, pre-GC and memory B cellsof pSS, Bm1, naïve, pre-GC and memory B cells of HC),respectively, we removed genes with coefficients of variationgreater than 0.3 according to each subpopulation after intensityfiltering and applied them to the WGCNA R package [27] followingthe package's tutorial. Briefly, we applied step-by-stepconstruction of the gene network and identification of modules tofileted gene expression dataset. First, we chose the propersoft-thresholding power based on the criterion of approximatescale-free topology. After calculating the adjacencies withselected soft-

thresholding power, we transformed the adjacency intoTopological Overlap Matrix, and calculate the correspondingdissimilarity. Then, we used hierarchical clustering to produce ahierarchical clustering tree of genes. Module identificationamounts to the identification of individual branches; we cut thebranches off the dendrogram with setting minimum module size 30. Asa result, 6 and 5 modules were generated in B cells of patientswith pSS and HC, respectively. Further, to annotate and estimatethe regulatory mechanism of co-expressed gene network, weinvestigated pathways, upstream regulators and functions associatedwith modules by using Ingenuity Pathway Analysis (IPA).

Cell culture and real-time quantitative PCR (qPCR)

B cell lines (Ramos, CCRF, U266B1) (each 1.0 × 10^5/mL) in RPMI 1640 medium (American Type Culture Collection)supplemented with 10% fetal calf serum were treated with differentconcentrations of IFNα, IFNγ and TNFα for 48 h. We next isolatedtotal RNA that was transcribed into cDNAs using NucleoSpin RNA(Macherey Nagel) and ReverTra Ace qPCR RT Master Mix (Toyobo). qPCRwas performed using PowerUp SYBR Green Master Mix (Thermo FisherScientific). The levels of total cellular RNA were similar amongthe cell lines. At least two biological repeats were performed foreach experiment. The levels of *GAPDH* mRNA were used toadjust the values of target transcripts determined using qPCRanalyses. We validated gene expression levels in $CD19^+$ Bcells of patients with pSS (n=14) and HC (n=12) (SupplementaryTable 2) using quantitative PCR (qPCR) analysis. Sequence ofprimers for qPCR analysis was described in Supplementary Table3.

Statistics

Continuous values are shown as the median and range of values.Differences between the groups were analysed using the Mann–Whitneytest for continuous variables. Correlations were analysed usingSpearman's correlation coefficient, unless otherwise noted. $P<0.05$ indicates a significant difference. Statistical tests wereconducted using R software version 3.5.2.

Results

Identification of gene signatures in B cell subsetsof pSS

First, we visualized the gene expression profile after intensityfiltration. Using principal component and hierarchical clusteringanalyses, we found that clinical status, namely pSS or HC, had lessinfluence on phenotype than the type of B cell subset (Figure 1Band C). Memory B cells, in particular, had a distinct expressionpattern compared with those of other cell types.

Expression of LINC00487 correlates with IFN signature genesand disease activity score of pSS

Using validation cohort, we tried to conduct qPCR analysis usingeach B cell subset. However, because *LINC00487* expressedat very low level in HCs, we couldn't sort enough each B cellsubset from them to detect expression of *LINC00487*.Therefore, considering that upregulation of *LINC00487* wascommon in

all B cell subsets in microarray cohort, we used a bulkof $CD19^+$ B cells in validation cohort. As a result, theexpression of *LINC00487* in $CD19^+$ B cells ofpatients with pSS was significantly upregulated compared with HC (p= 0.035) (Figure 3D). As the same as Figure 3A and B, expression of*LINC00487* was correlated with *IFI44L* with lowerp-value in patients with pSS ($p < 0.001$) (Figure 3E) comparedwith that of HC (p=0.04) (Figure 3F). To confirm the induction of*LINC00487* in response to IFNα, we used three types of Bcell lines and measured *LINC00487* expression using qPCR.As with *IFI44L*, *LINC00487* was strongly induced bytreatment with IFNα (Figure 3E and F), unlike TNFα and IFNγ(Supplementary Figure 3A), suggesting that IFNα is an upstreamregulator of *LINC00487*. Compared with clinical features,the expression of *LINC00487* tended to be higher inpatients with higher serum IgG levels and lower lymphocyte countswhich were characteristic for pSS (Supplementary Figure 3B).

Co-expression gene networks constructed for each B cellsubset

To identify transcriptional networks in the B cells of patientswith pSS, we performed WGCNA. WGCNA can identify novel geneinteractions which might be missed in the DEGs analysis, becauseWGCNA is based on correlations of gene expression [25].

After selection, we identified a new set comprising 3634 genes(Supplementary Table 6-7). In principal component analysis usingthese gene sets, Bm1 and naive B cells in pSS tended to be closerto the expression patterns of a more highly differentiated subset,respectively, compared to those in HC; i.e., Bm1 to naive and naiveto pre-GC (Figure 4A). WGCNA identified 6 modules in pSS and anassociation between cell subsets and modules (Figure 4B). Forexample, the yellow and brown modules are associated with thepre-GC B cell subset, whereas the turquoise module is associatedwith the memory B cell subset. The grey module of pSS was enrichedin Bm1/naïve B cell subsets with various ranges of eigengenes. Inthe analysis to detect an association between modules and clinicaltraits, the grey module significantly correlated with the clinicaldisease activity score (Figure 4C). Similarly, another five moduleswere identified in HC (Supplementary Figure 4A). The module pairswith the greatest number of genes in each pSS module overlappingwith either HC modules were blue (pSS) – turquoise (HC), brown(pSS) – brown (HC), grey (pSS) – blue (HC), green (pSS) – turquoise(HC), yellow (pSS) – blue (HC), and turquoise (pSS) – turquoise(HC), respectively (Supplementary Figure 5). All of the above pairsoverlapped more than half of the gene list, suggesting that theoverall picture of the B cell modules in pSS and the B cell modulesin HC were similar.

To extract the characteristics of pSS modules, we performedfunctional analysis using Ingenuity Pathway Analysis (IPA). Thecluster of enriched pathway, upstream regulator anddisease/biological functions in modules of pSS were distinguishedfrom those of HC (Figure 5A and Supplementary Tables 8–13). The topfive most significantly enriched pathways, regulators anddisease/biological functions in modules of pSS are shown in Figure5B. In upstream regulator analysis, transcriptional regulators suchas Sry-related high mobility group box (Sox) family such as*SOX4* were identified in the yellow and blue modules ofpSS. Focusing on specificity (Supplementary Figure 6 andSupplementary Table 14-16), mir-21-5p was identified as the topsignificant regulator in the yellow module of pSS. To search forhub genes, we

described a co-expressed gene network using tophighly interconnected genes defined by topological overlap in eachmodule (Figure 5C). Several genes bridged inter-modules, namely hubgenes, such as *SOX4*. *SOX4* was not identified inthe co-expression network of HC (Supplementary Figure 4B),suggesting its involvement in the dysregulation of B cells inpSS.

Discussion

In the present study, we conducted transcriptome analysis of Bcell subpopulations. First, we found that expression of*LINC00487* was upregulated in B cell subsets derived frompatients with pSS. Further, *LINC00487* expressionsignificantly correlated with disease activity and was induced byIFNα stimulation. Second, using WGCNA, we identified several keynetworks and hub genes in the B cells of patients with pSS.

IFN signalling is a central component of the pathogenesis of pSS[28]. To our knowledge, this is the first study to showupregulation of IFN signalling in B cells of patients with pSSaccording to developmental subsets. Type 1 IFN signalling plays acrucial role in the development of autoreactive B cells in a mousemodel of autoimmune disease [5]. Principal component analysis usinggene sets for WGCNA demonstrated that Bm1/naïve B cells of pSS wereclearly distinguished from those of HC and showed a trend ofexpression patterns closer to the more highly differentiatedsubset. These results suggest that molecular dysfunction, such asdisruption of peripheral tolerance, might start at an early stagein the maturation of B cells. Indeed, the frequencies of naïve Bcells expressing autoreactive antibodies are significantlyincreased in patients with pSS [29].

Interestingly, expression of the HLA class II gene wasupregulated. GWAS studies of pSS found associations of*HLA-DQA1* and *HLA-DQB1* loci [6,7]. Further, IFNαinduces the expression of *HLA-DQA1* [30]. In patients withpSS, *HLA-DQA1* and *HLA-DQB1*alleles are associated with higher concentrations of anti-SSA andSSB antibodies [31]. These findings suggest that aberrantinteractions among IFN signalling and HLA class II genes maytrigger the breakdown of B cell tolerance, leading to thedevelopment of pSS.

LINC00487 was upregulated in all B cell subsets of pSSand its expression significantly correlated with the diseaseactivity scores of pSS and ISGs, although ESSDAI score of ourpatients skewed toward the low end. Moreover, we found that IFNαwas an upstream regulator of *LINC00487* in B cells. Thesequence of *LINC00487*, which belongs to the class of longintergenic non-coding RNAs and resides on human chromosome 2, isatypically long (> 40,000 bases). Although *LINC00487* isone of the hub genes in the normal development of human B cells[32], many of its properties, including its function, areunexplained. Further, there is no ortholog or paralog of this genein species other than humans that may provide clues to its functionin human cells. However, according to AceView, one of threetranscriptional variants derived from *LINC00487* haspotential to encode a protein in silico prediction [33].

Four genomic locations are considered candidate enhancers of*LINC00487* transcription. The targets of the regulators of*LINC00487* overlap with those of other ISGs [34]. Moreover,referring to the public transcriptome database of microarrayanalyses of healthy humans, expression of *LINC00487* ishigher

specifically in centroblasts and centrocytes of the germinalcentre [35]. IFNα promotes the autoreactivity of B cells viagerminal centre pathways [5]. Further, *LINC00487*expression is upregulated in the subgroup of diffuse large B-celllymphoma with molecular characteristics of germinal centre B cells,compared with other subgroups, and is associated with the efficacyof B cell depletion therapy [36]. Therefore, our study suggeststhat upregulation of *LINC00487* expression in all B cellsubsets may reflect or regulate the enrichment of a germinalcentre-like reaction by IFNα from an early stage of B celldevelopment, leading to B cell autoreactivity in patients withpSS.

Gene co-expression analysis revealed an aberrant network in Bcell subpopulations. The top significant upstream regulator of thegrey module of pSS, which was associated with clinical diseaseactivity score and enriched in the early stage of B celldevelopment in pSS, was the gene encoding T-cell acute lymphocyticleukemia protein 1 (*TAL1*). B cell development isstringently controlled by stage-specific transcription factors. Thetranscription factor *TAL1* regulates genes such as IKAROSfamily zinc finger 3 (*IKZF3*), which is one of hub genes inthe grey module of pSS (Figure 5C). *IKZF3* is alineage-specific transcription factor that is important in theregulation of B cell proliferation and development [37].

In the pre-GC B cells-associated module of pSS, *SOX4*was identified as an upstream regulator and a hub gene. In mice,*SOX4* regulates the differentiation of early-stage B cellsby activating the expression of *Rag1* and *Rag2*[38]. Further, *SOX4* contributes to the formation ofectopic lymphoid-like structures via promoting CXCL13, which isligand of CXCR5 on naïve B cells and critical for migration intolight zone of germinal centre undergoing somatic hypermutation,production from$PD\text{-}1^{hi}CXCR5^{-}CD4^{+}$ T cells[39]. Additionally, in proteome analysis, CXCL13 positivelycorrelates with the disease activity score and serum IgG levels ofpatients with pSS [19]. Further study is needed to explore the roleof *SOX4* in mature B cells.

Further, we identified miR-21 as an upstream regulator of theyellow module of pSS. Although miR-21 is upregulated in peripheralblood mononuclear cells of pSS [40], the present study is the firstto propose its involvement in pSS via B cell dysregulation.Interestingly, miR-21 regulates the immune response of memory Tcells via induction of transcription networks, such as*SOX4* [41], implying that the interaction between miR-21and *SOX4* may also affect function of B cells.

pSS is enormously heterogeneous disease from the molecular pointof view. In WGCNA, despite filtering out genes with high variationin each cell subpopulation, module expression showed the strongvariation among samples in the same group, in particular Bm1 subset(Figure 4B). This finding suggested that there were differentregulatory mechanisms within each subgroup, although it wasinformation that couldn't be analyzed due to the limited samplesize.

The current study suffers from several limitations. First,patients in our cohort have low disease activity/severity. Becausethere were few patients with high disease activity beforeimmunosuppressive treatment, it was difficult to include suchpatients in this study. Therefore, correlation between expressionof *LINC00487* and disease activity score is needed to bevalidated in other cohort including patients with high ESSDAIscores. Second, healthy controls were younger than the pSSpatients, and this

could be a confounding variable. Third, samplesize was limited. To overcome limitation about small sample size,we validated by qPCR using another cohort, supporting resultsderived from microarray analysis. However, regarding WGCNA, wecouldn't include subjects enough to replicate by qPCR.

Conclusions

Our focus on B cell subpopulation using a multi-level approachemploying analysis of DEGs and WGCNA identified significant genesand networks as novel players in the pathogenesis of pSS. Toconfirm our results, further study is needed.

Declarations

Ethics approval and consent to participate

Ethics approval was obtained from the Institutional Review Boardof Keio University School of Medicine (IRB No. 20110258). Consentto participate was obtained from all subjects in the current studybefore blood specimen was collected.

Consent for publication

Not applicable.

Availability of data and materials

The transcriptome data are available at the GEO database. Theaccession codes are GSE135809. All custom computer codes in thegeneration or processing of the described data are available uponreasonable request. Supplementary Table 4-18 and R script for WGCNAare deposited in figshare (https://figshare.com/articles/Supplementary_Table_4-14/11683959).

Competing interests

YK, YO, MTaki and RK are employed by Takeda PharmaceuticalCompany Limited. S.T. was employed by Takeda Pharmaceutical CompanyLimited. KS has received research grants from Eisai, Bristol-MyersSquibb, Kissei Pharmaceutical, and Daiichi Sankyo, and speakingfees from Abbie Japan, Astellas Pharma, Bristol-Myers Squibb,Chugai Pharmaceutical, Eisai, Fuji Film Limited, JanssenPharmaceutical, Kissei Pharmaceutical, Mitsubishi TanabePharmaceutical, Pfizer Japan, Shionogi, Takeda Pharmaceutical, andUCB Japan, consulting fees from Abbie, and Pfizer Japan. AY hasreceived speaking fees from Chugai Pharmaceutical, MitsubishiTanabe Pharmaceutical, Pfizer Japan, Ono Pharmaceutical, Maruho,and Novartis, and consulting fees from GSK Japan. TT has receivedresearch grants from Astellas Pharma Inc, Bristol-Myers KK, ChugaiPharmaceutical Co. Ltd., Daiichi Sankyo Co. Ltd, TakedaPharmaceutical Co. Ltd, Teijin Pharma Ltd, AbbVie GK, AsahikaseiPharma Corp, Mitsubishi Tanabe Pharma Co, Pfizer Japan Inc, andTaisho Toyama Pharmaceutical Co. Ltd, Eisai Co. Ltd, AYUMIPharmaceutical Corporation, and Nipponkayaku Co. Ltd, and speakingfees from AbbVie GK,

Bristol-Myers KK, Chugai Pharmaceutical Co.Ltd, Mitsubishi Tanabe Pharma Co, Pfizer Japan Inc, and AstellasPharma Inc, and Diaichi Sankyo Co. Ltd, and consultant fees fromAstra Zeneca KK, Eli Lilly Japan KK, Novartis Pharma KK, MitsubishiTanabe Pharma Co, Abbivie GK, Nipponkayaku Co. Ltd, JanssenPharmaceutical KK, Astellas Pharma Inc, and Taiho PharmaceuticalCo. Ltd. JI and MT declare no potential conflict of interest.

Funding

This study was funded by Keio University (grunt number:04-013-0188) and Takeda Pharmaceutical Company Limited (gruntnumber: 04-078-0067).

Authors' Contributions

Conceptualization: JI, KS, MTake, YK, YO, MTaki, RK and ST.Funding acquisition: KS and TT. Data acquisition: YK, MTaki and RK.Formal analysis: JI, YO and ST. Supervision: KS. Writing andoriginal draft preparation: JI. Writing review and editing: KS, AYand TT.

Acknowledgements

This work was supported by Takeda Pharmaceutical CompanyLimited, Kanagawa, Japan (grant number 04-078-0067) and KeioUniversity (grant number 04-013-0188).

We thank Ms. Harumi Kondo, Ms. Mayumi Ota, Ms. Yoshiko Yogiashi,Ms. Yuki Otomo, Mr. Fumitsugu Yamane for helping with theexperiments, and Digital Medical Communications Corp. forproofreading sentences for English proofreading.

Abbreviations

DZ, dark zone: ESSDAI, EULAR Sjögren's Syndrome Disease ActivityIndex: FACS, fluorescence-activated cell sorting: GC-B, germinalcentre B cell: HC, healthy control: HLA, human leukocyte antigen:IFN, interferon: IKZF3, IKAROS family zinc finger 3: ISGs,interferon-stimulated genes: LZ, light zone: lncRNA, longnon-coding RNA: pSS, primary Sjögren's syndrome: TAL1, T-cell acutelymphocytic leukemia protein 1: WGNCA, weighted gene co-expressionnetwork analysis.

References

1. Qin B, Wang J, Yang Z, et al. Epidemiology of primary Sjögren'ssyndrome: a systematic review and meta-analysis. Ann Rheum Dis.2015;74:1983-1989.
2. Muskardin TLW, Niewold TB. Type I interferon in rheumaticdiseases. Nat Rev Rheumatol 2018;14:214-228.
3. Imgenberg-Kreuz J, Sandling JK, Björk A, et al. Transcriptionprofiling of peripheral B cells in antibody-positive primarySjögren's syndrome reveals upregulated expression of CX3CR1 and atype I and type

II interferon signature. Scand J Immunol.2018;87:e12662.

4. Devauchelle-Pensec V, Cagnard N, Pers JO, et al. Geneexpression profile in the salivary glands of primary Sjögren'ssyndrome patients before and after treatment with rituximab.Arthritis Rheum. 2010;62:2262-71.
5. Domeier PP, Chodisetti SB, Schell SL, et al. B-Cell-IntrinsicType 1 Interferon Signaling Is Crucial for Loss of Tolerance andthe Development of Autoreactive B Cells. Cell Rep.2018;24:406-418.
6. Lessard CJ, Li H, Adrianto I, et al. Variants at multiple lociimplicated in both innate and adaptive immune responses areassociated with Sjögren's syndrome. Nat Genet 2013;45:1284–92.
7. Li Y, Zhang K, Chen H, et al. A genome-wide association studyin Han Chinese identifies a susceptibility locus for primarySjögren's syndrome at 7q11.23. Nat Genet 2013;45:1361-5.
8. Triantafyllopoulou A, Moutsopoulos H. Persistent viralinfection in primary Sjögren’s syndrome: review and perspectives.Clin Rev Allergy Immunol. 2007;32:210-214.
9. Reale M, D'Angelo C, Costantini E, et al. MicroRNA in Sjögren'sSyndrome: Their Potential Roles in Pathogenesis and Diagnosis. JImmunol Res. 2018:7510174.
10. Wang-Renault SF, Boudaoud S, Nocturne G, et al. Deregulation ofmicroRNA expression in purified T and B lymphocytes from patientswith primary Sjögren's syndrome. Ann Rheum Dis2018;77:133–140.
11. Tang Y, Zhou T, Yu X. The role of long non-coding RNAs inrheumatic diseases. Nat Rev Rheumatol. 2017;13:657-669.
12. Sandhya P, Joshi K, Scaria V. Long noncoding RNAs could bepotential key players in the pathophysiology of Sjögren's syndrome.Int J Rheum Dis. 2015; 18:898-905.
13. Shi H, Cao N, Pu Y, et al. Long non-coding RNA expressionprofile in minor salivary gland of primary Sjögren's syndrome.Arthritis Res Ther. 2016;18:109.
14. Tasaki S, Suzuki K, Nishikawa A, et al. Multiomic diseasesignatures converge to cytotoxic CD8 T cells in primary Sjögren’ssyndrome. Ann Rheum Dis 2017;76:1458–1466.
15. GTEx Consortium. Human genomics. The Genotype-Tissue Expression(GTEx) pilot analysis: multitissue gene regulation in humans.Science. 2015;348:648-660.
16. Nishikawa A, Suzuki K, Kassai Y, et al. Identification ofdefinitive serum biomarkers associated with disease activity inprimary Sjögren's syndrome. Arthritis Res Ther. 2016;18:106.
17. Lee SJ, Oh HJ, Choi BY, et al. Serum 25-Hydroxyvitamin D3 andBAFF Levels Are Associated with Disease Activity in PrimarySjogren's Syndrome. J Immunol Res. 2016;2016:5781070.
18. James K, Chipeta C, Parker A, et al. B-cell activity markersare associated with different disease activity domains in primarySjögren's syndrome. Rheumatology (Oxford). 2018;57:1222-1227.
19. Ishioka-Takei E, Yoshimoto K, Suzuki K, et al. Increasedproportion of a CD38highIgD+ B cell subset in peripheral blood isassociated with clinical and immunological features in patientswith primary Sjögren's syndrome. Clin Immunol. 2018;187:85-91.
20. Bohnhorst JØ, Bjørgan MB, Thoen JE, et al. Bm1-Bm5classification of peripheral blood B cells reveals circulatinggerminal center founder cells in healthy individuals anddisturbance in the B cell

subpopulations in patients with primarySjögren's syndrome. J Clin Invest. 2005;115:3205-3216.

21. Vitali, S. Bombardieri, R. Jonsson, et al. Classificationcriteria for Sjögren's syndrome: a revised version of the Europeancriteria proposed by the American-European Consensus Group. AnnRheum Dis. 2002;61:554–558.
22. C. Shiboski, C.H. Shiboski, L. Criswell, et al. AmericanCollege of Rheumatology classification criteria for Sjögren'ssyndrome: A data-driven, expert consensus approach in the Sjögren'sInternational Collaborative Clinical Alliance Cohort. ArthritisCare Res (Hoboken). 2012;64:475-87.
23. Seror R, Ravaud P, Bowman SJ, et al. EULAR Sjögren's syndromedisease activity index: development of a consensus systemic diseaseactivity index for primary Sjögren's syndrome. Ann Rheum Dis.2010;69:1103-9.
24. Seror R, Bowman SJ, Brito-Zeron P, et al. EULAR Sjögren'ssyndrome disease activity index (ESSDAI): a user guide. RMD Open.2015;1:e000022.
25. Chen EY, Tan CM, Kou Y, et al. Enrichr: interactive andcollaborative HTML5 gene list enrichment analysis tool. BMCBioinformatics. 2013;128.
26. Huang R, Grishagin I, Wang Y, et al. The NCATS BioPlanet - AnIntegrated Platform for Exploring the Universe of CellularSignaling Pathways for Toxicology, Systems Biology, and ChemicalGenomics. Front Pharmacol. 2019;10:445.
27. Langfelder P, Horvath S. WGCNA: an R package for weightedcorrelation network analysis. BMC Bioinformatics. 2008;9:559.
28. Nocturne G, Mariette X. B cells in the pathogenesis of primarySjögren’s syndrome. Nat Rev Rheumatol 2018;14:133-145.
29. Glauzy S, Sng J, Bannock JM, et al. Defective Early B CellTolerance Checkpoints in Sjögren's Syndrome Patients. ArthritisRheumatol. 2017;69:2203-2208.
30. Harris PE, Malanga D, Liu Z, et al. Effect of interferon alphaon MHC class II gene expression in ex vivo human islet tissue.Biochim Biophys Acta. 2006 ;1762:627-635.
31. Harley JB, Reichlin M, Arnett FC, et al. Gene interaction atHLA-DQ enhances autoantibody production in primary Sjögren'ssyndrome. Science. 1986;232:1145-7.
32. Petri A, Dybkær K, Bøgsted M, et al. Long Noncoding RNAExpression during Human B-Cell Development. PLoS One. 2015;10:e0138236.
33. Thierry-Mieg D, Thierry-Mieg J. AceView: a comprehensivecDNA-supported gene and transcripts annotation. Genome Biol. 2006;7Suppl 1:S12.1-14.
34. Fishilevich S, Nudel R, Rappaport N, et al. GeneHancer:genome-wide integration of enhancers and target genes in GeneCards.Database (Oxford). 2017;2017:bax028.
35. McCall MN, Uppal K, Jaffee HA, et al. The Gene ExpressionBarcode: leveraging public data repositories to begin catalogingthe human and murine transcriptomes. Nucleic Acids Res.2011;39:D1011-5.

36. Jiayu Yu, Alyssa Bouska, Waseem Lone, et al. Long Non CodingRNA Signatures with Diagnostic and Prognostic Significance inAggressive B-Cell Lymphoma. Blood 2016;128:1759.
37. Thompson EC, Cobb BS, Sabbattini P, et al. Ikaros DNA-bindingproteins as integral components of B celldevelopmental-stage-specific regulatory circuits. Immunity.2007;26:335-44.
38. Mallampati S, Sun B, Lu Y, et al. Integrated genetic approachesidentify the molecular mechanisms of Sox4 in early B-celldevelopment: intricate roles for RAG1/2 and CK1ε. Blood.2014;123:4064-76.
39. Yoshitomi H, Kobayashi S, Miyagawa-Hayashino A, et al. HumanSox4 facilitates the development of CXCL13-producing helper T cellsin inflammatory environments. Nat Commun. 2018;9:3762.
40. Chen JQ, Papp G, Póliska S, et al. MicroRNA expression profilesidentify disease-specific alterations in systemic lupuserythematosus and primary Sjögren's syndrome. PLoS One.2017;12:e0174585.
41. Kim C, Hu B, Jadhav RR, et al. Activation of miR-21-RegulatedPathways in Immune Aging Selects against Signatures Characteristicof Memory T Cells. Cell Rep. 2018;25:2148-2162.e5.

Figures

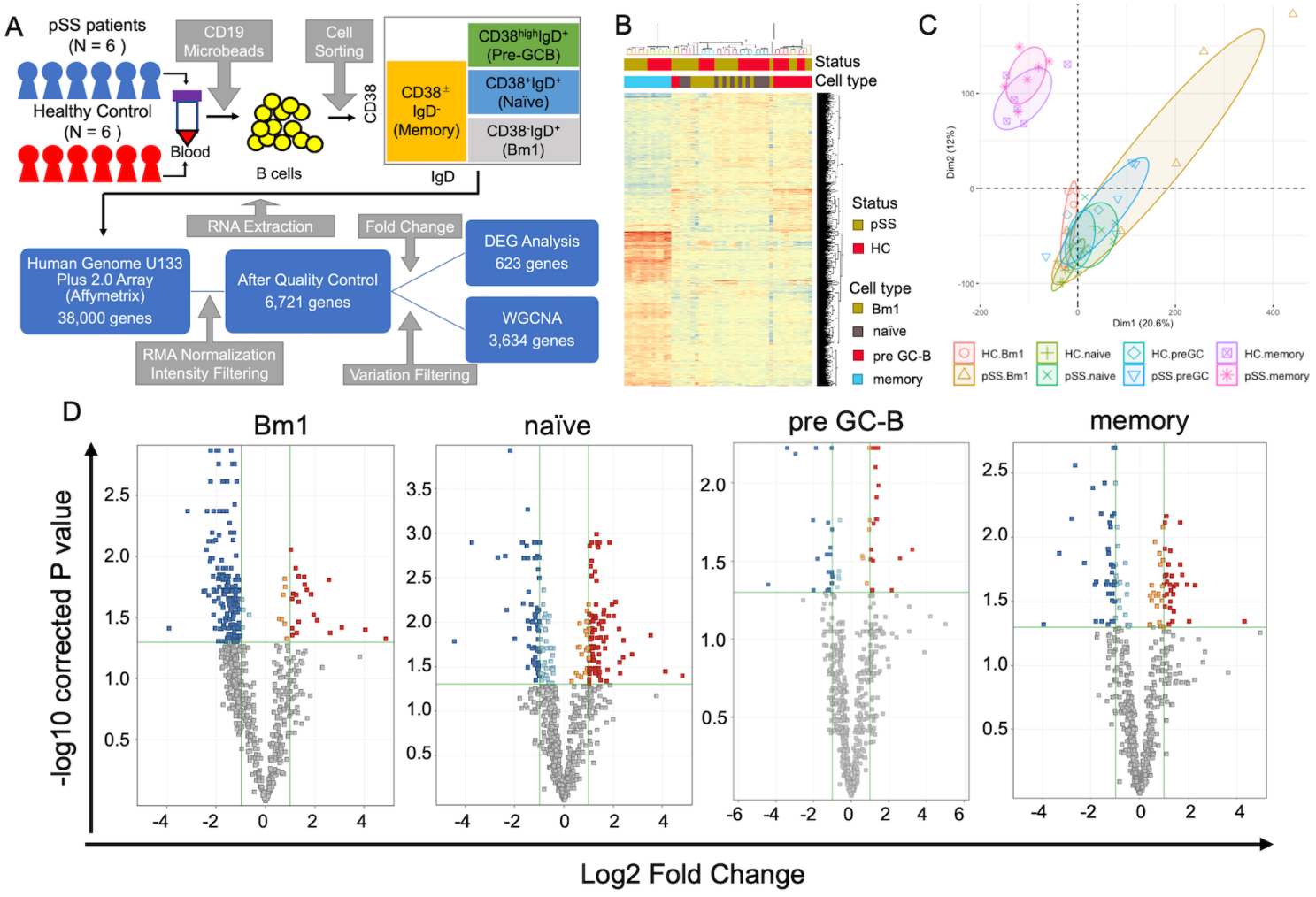


Figure 1

Gene expression profile of B cell subsets in patients with pSS and HC. (A) Study protocol. (B) Hierarchical clustering analysis and (C) principal component analysis to summarize the dissimilarity of the transcriptome of each B cell subset. Rows correspond to genes and columns to samples. The ellipse shows the 50% confidence interval of the value of principal component analysis. (D) Dysregulated genes are shown in the volcano plot. Horizontal green lines show the cut-off of the p-value indicating a significant difference, and the vertical green lines show a log2-fold change. DEG, differentially expressed gene; GC-B, germinal centre B cell: HC, healthy control: pSS, primary Sjögren's syndrome; WGCNA, weighted gene co-expression network analysis.

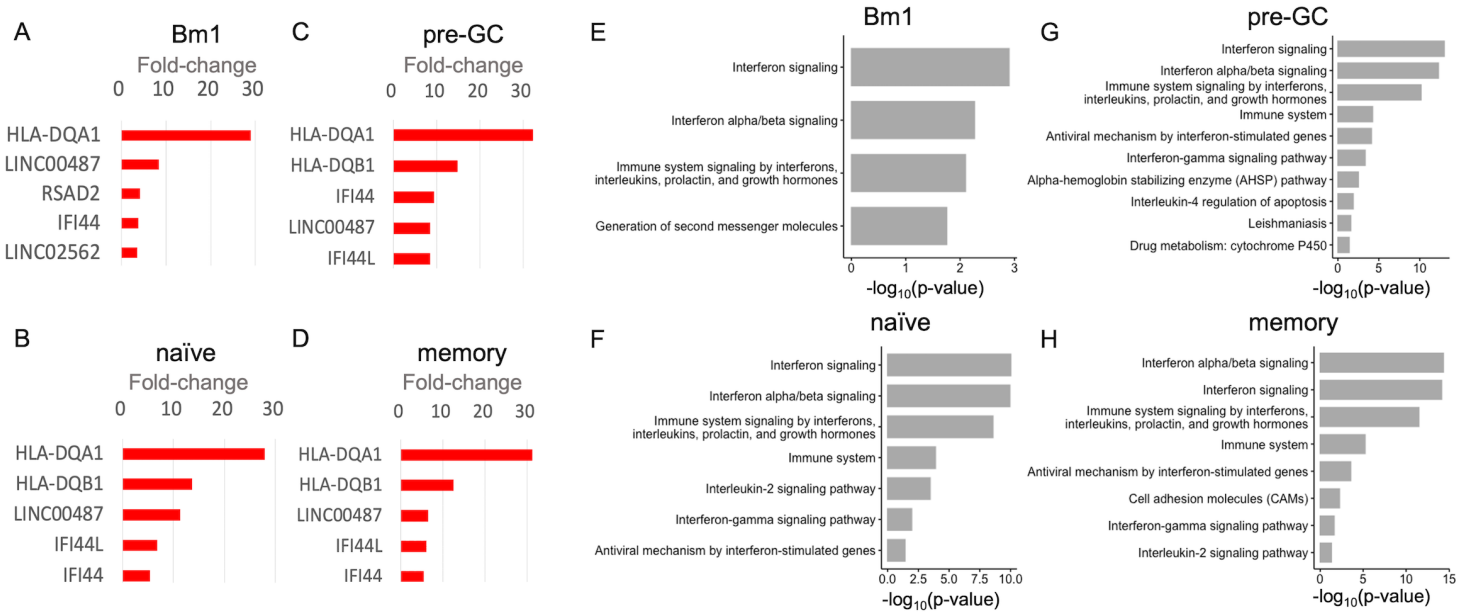


Figure 2

Gene signatures of B cell subsets of patients with primary Sjögren's syndrome. (A-D) The top upregulated five genes in Bm1 (A), naïve (B), pre-GC (C), and memory (D) B cell subset of patients with pSS compared with those of the HC. (E-H) Upregulated pathways in Bm1 (E), naïve (F), pre-GC (G), and memory (H) B cell subset of patients with pSS compared with those of the HC. GC, germinal centre.

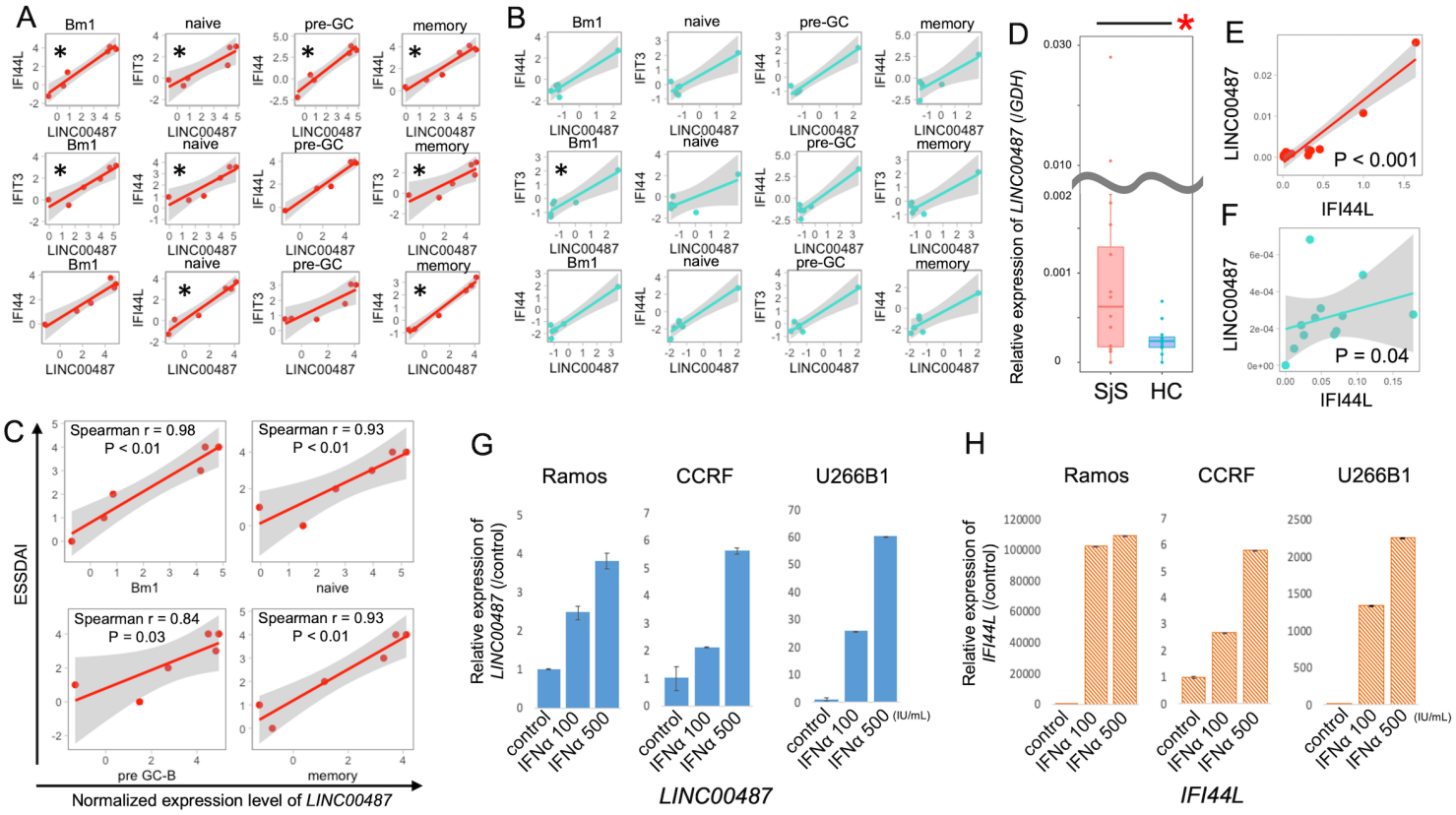


Figure 3

Characteristics of LINC00487. (A-B) Correlation between LINC00487 and interferon signature genes (IFI44L, IFIT3, IFI44) of pSS (A) and HC (B) according to subset. Significant p-value was shown in each plot. '*', p<0.05; Spearman's correlation test. (C) Scatter plot of disease activity scores and normalized expression levels of LINC00487 in Bm1, naïve, pre-GC and memory subset of patients with pSS. (D-F) qPCR analysis of LINC00487 in primary human CD19+ B cells. Comparison of the relative expression of LINC00487 (/GDH) between patients with pSS and HC (D) ('*', p<0.05; the Mann-Whitney test), and correlation plot of relative expression of the LINC00487 (/GDH) and interferon induced protein 44-like gene (/GDH) in patients with pSS (E) and HC (F). (G-H) qPCR analysis of B cell lines after treatment with IFNα. B cell lines were treated for 48 h. Results are represented as the mean ± standard deviation. ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index: GC-B, germinal centre B cell: GDH, glyceraldehyde 3-phosphate dehydrogenase; HC, healthy control: IFI44, interferon induced protein 44; IFI44L, interferon induced protein 44 like; IFIT3, interferon induced protein with tetratricopeptide repeats 3; pSS, primary Sjögren's syndrome.

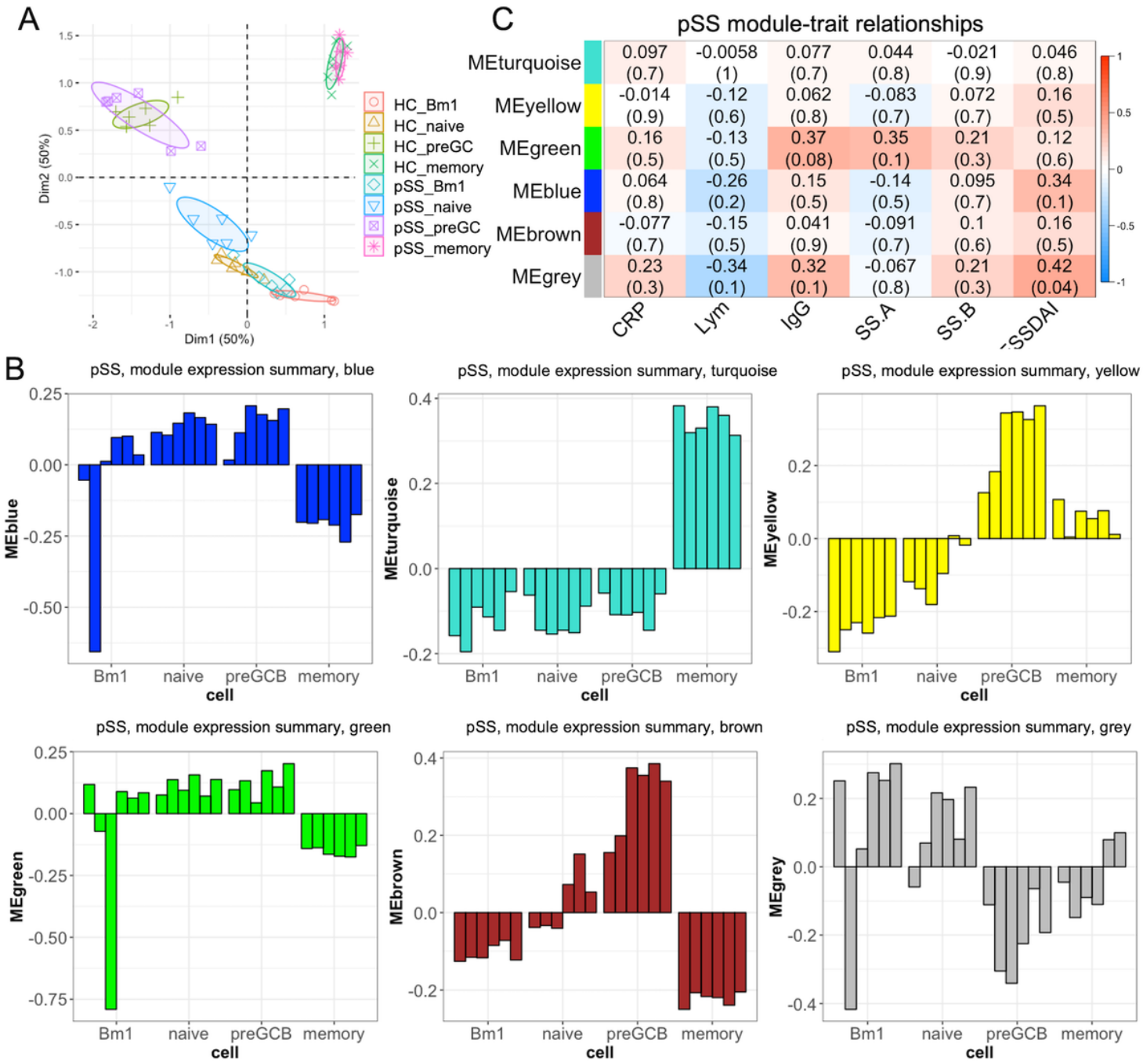
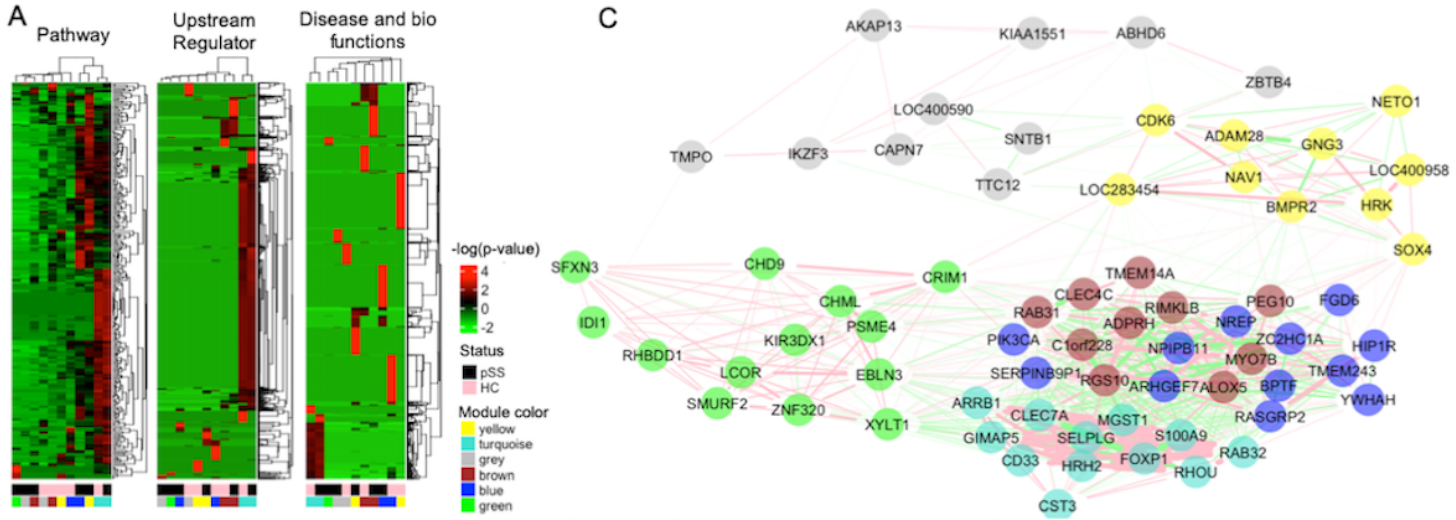


Figure 4

Weighted gene co-expression network analysis of B cell subsets of patients with pSS. (A) Principal component analysis using the gene set produced generated using WGCNA. The ellipse shows the 50% confidence interval of the value of principal component analysis. (B) Module eigengene of each B cell subset of patients with pSS. The y-axis displays the values of the module eigengene, and the x-axis displays each cell subset. (C) The heat map shows module-trait associations. Each row corresponds to a module's first eigengene and columns to a value of trait. Each cell contains the corresponding Spearman's correlation coefficient (p-value) between a value of each trait and a module's first eigengene. CRP, C-reactive protein: ESSDAI, EULAR Sjögren's Syndrome Disease Activity Index: GC-B, germinal centre B cell:

Lym, lymphocyte count: SS.A, anti-Sjögren's syndrome-related antigen A: ME, module eigengene: SS.B, anti-Sjögren's syndrome-related antigen B.



B

Pathway

yellow	
Ingenuity Canonical Pathways	-log(p-value)
Tec Kinase Signaling	5.16
Prolactin Signaling	4.74
Protein Kinase A Signaling	4.74
Molecular Mechanisms of Cancer	4.67
FAK Signaling	4.36

turquoise	
Ingenuity Canonical Pathways	-log(p-value)
Oxidative Phosphorylation	14.6
Mitochondrial Dysfunction	11.3
B Cell Development	7.20
Phagosome Formation	6.36
Germ Cell-Sertoli Cell Junction Signaling	5.59

grey	
Ingenuity Canonical Pathways	-log(p-value)
Cleavage and Polyadenylation of Pre-mRNA	1.53
Tight Junction Signaling	1.19
nNOS Signaling in Skeletal Muscle Cells	1.02
Triacylglycerol Degradation	1.01
Role of Oct4 in Mammalian Embryonic Stem Cell Pluripotency	0.97

brown	
Ingenuity Canonical Pathways	-log(p-value)
Wnt/β-catenin Signaling	1.96
Amyloid Processing	1.56
Epithelial Adherens Junction Signaling	1.41
Actin Nucleation by ARP-WASP Complex	1.39
Ovarian Cancer Signaling	1.38

blue	
Ingenuity Canonical Pathways	-log(p-value)
Prolactin Signaling	3.53
IGF-1 Signaling	3.09
UVB-Induced MAPK Signaling	3.04
Endometrial Cancer Signaling	2.93
Apelin Endothelial Signaling Pathway	2.91

green	
Ingenuity Canonical Pathways	-log(p-value)
Chondroitin Sulfate Biosynthesis	2.32
Dermatan Sulfate Biosynthesis	2.29
Diphthamide Biosynthesis	2.23
Heparan Sulfate Biosynthesis	2.13
Trans, trans-farnesyl Diphosphate Biosynthesis	2.10

Upstream Regulator

yellow	
Upstream Regulator	-log(p-value)
SOX4	5.16
SOX11	4.03
miR-21-5p	3.40
IL2	3.22
beta-estradiol	3.05

turquoise	
Upstream Regulator	-log(p-value)
lipopolysaccharide	26.68
IL15	25.80
IL5	25.41
TP53	24.41
Immunoglobulin	23.13

grey	
Upstream Regulator	-log(p-value)
TAL1	2.79
oligomycin A	2.62
allose	2.62
MLXIP	2.32
iberdomide	2.32

brown	
Upstream Regulator	-log(p-value)
methylselenic acid	5.24
IL2	3.79
CXCR4	3.67
miR-873-5p	3.57
MAP3K14	3.56

blue	
Upstream Regulator	-log(p-value)
SOX11	5.02
SOX4	3.15
BRD1	2.93
PRDM5	2.83
RCE1	2.68

green	
Upstream Regulator	-log(p-value)
NTNG2	2.71
miR-615-3p	2.24
HIC1	2.21
PIP5K1A	2.11
VLDL	2.11

Disease and bio functions

yellow	
Diseases or Functions Annotation	-log(p-value)
Cell death of immune cells	4.24
Quantity of leukocytes	4.22
Quantity of mononuclear leukocytes	3.97
Binding of phosphatidylinositol 3,5-diphosphate	3.97
Quantity of blood cells	3.96

turquoise	
Diseases or Functions Annotation	-log(p-value)
Cell death of blood cells	25.94
Cell death of immune cells	24.58
Systemic autoimmune syndrome	21.45
Degranulation of neutrophils	21.36
Proliferation of mononuclear leukocytes	21.34

grey	
Diseases or Functions Annotation	-log(p-value)
Autosomal recessive congenital cataract	2.94
Egression of naive T lymphocytes	2.60
Arrest in Gap 0-Gap 1 phase of kidney cell lines	2.60
Microcephaly, mental retardation, and distinctive facies with cardiac and genitourinary malformations	2.60
Hydrolysis of lysobisphosphatidic acid	2.60

brown	
Diseases or Functions Annotation	-log(p-value)
Secondary tumor	4.73
Growth of malignant tumor	4.40
Growth of tumor	4.40
Morphology of vascular sprout	4.10
Degranulation of phagocytes	4.04

blue	
Diseases or Functions Annotation	-log(p-value)
Targeting of cellular membrane	3.91
Infiltration by myofibroblasts	3.91
Abnormal morphology of pre-B lymphocytes	3.38
Formation of dendritic spines	3.28
Abnormal morphology of transitional B lymphocytes	3.22

green	
Diseases or Functions Annotation	-log(p-value)
Synthesis of chondroitin sulfate	2.82
Entry into cell cycle progression of peripheral T lymphocyte	2.70
Desbuquois syndrome type 2	2.70
Endocrine-cerebroosteodysplasia	2.70
Neurodevelopmental disorder with dysmorphic facies and distal limb anomalies	2.70

Figure 5

Gene network characteristics and hub genes in the gene networks of B cells of pSS. (A) Hierarchical clustering analysis with heatmap of enriched pathways (left), upstream regulators (middle) and disease/functions (right). (B) Top five significantly enriched canonical pathways (left), upstream

regulators (middle) and disease & biological functions (right) in the modules of pSS. The colour of each frame corresponds to module-colour. (C) Associations of intra- and inter-modular hub genes in patients with pSS. Node-colour corresponds to module-colour. The pink-coded edge represents the correlation, and the green-coded edge represents an inverse correlation. The width of the edge reflects the absolute weight of a correlation.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement1.pdf